DOCUMENT RESUME

ED 433 819 IR 057 449

AUTHOR	Darzentas, Jenny
TITLE	Sharing Metadata: Enabling Online Information Provision.
PUB DATE	1999-05-00
NOTE	10p.; In: The Future of Libraries in Human Communication: Abstracts and Fulltext Documents of Papers and Demos Given at the [International Association of Technological University Libraries] IATUL Conference (Chania, Greece, May 17-21, 1999) Volume 19; see IR 057 443.
AVAILABLE FROM	Web site: http://educate.lib.chalmers.se/IATUL/proceedcontents/chanpap/darzenta.html
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Libraries; Bibliographic Utilities; *Cataloging; *Cooperative Programs; Distance Education; *Electronic Libraries; Foreign Countries; Higher Education; Information Management; Instructional Materials; *Library Cooperation; Library Role; Library Services; *Shared Library Resources; Union Catalogs
IDENTIFIERS	*Electronic Resources; Greece; Learning Environments; MARC; Metadata; OCLC

ABSTRACT

This paper describes work being carried out in the fields of online education provision and library systems, beginning with a description of the current state of the art with regard to online learning environments and educational materials management. Suggestions and solutions for librarians dealing with the management of educational digital content for education service providers are presented, including cooperation and collaboration, linking publishers and national bibliographies, MARC and metadata, and union catalogs/virtual union catalogs. An overview is provided of UNIverse, a European-funded project that is developing a library system to support a virtual union catalog and that offers mechanisms for facilitating cataloging activities by enabling record supply. The paper then focuses on the experiences of a sub-group of the UNIverse project, headed by the National Library of Greece, to test and evaluate the record sharing capabilities of the system and collaborative cataloging in practice. (Contains 24 references.) (MES)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

N. Fjallbrant

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED 433 819

SHARING METADATA: ENABLING ON LINE INFORMATION PROVISION

Darzentas, Jenny
Kyros
55 Evripidou & Thisseos Av, Kallithea 176 74 Athens, Greece

E-mail: kyros@compulink.gr

Introduction

Online education, although by no means perfected, is now a reality. Hand in hand with its development are the continuing advances in education materials management. This paper describes work being carried out both in the field of online education provision and library systems. It briefly describes a prototype online learning environment (GESTALT) [1] and highlights the implications of such environments on libraries in terms of discovery of course components and relevant support material. The task of cataloguing, already one of the most heavy in terms of human resources, becomes an increased burden when it relates to digital material. It becomes necessary to describe not only the content and form and location of such material, but also, other metadata concerning its accessibility and delivery media. Again digital material may be composed of many separate components which each carry separate cataloguing requirements. In the context of the learning environment, a lecture may have text, sound, graphics, video, self-assessment exercises, a bibliography with hyperlinks. It is possible to tag all these digital objects with metadata in order to describe them and also to aggregate/desegregate so that the material may be used in a highly modular way.

Such a vision of online information provision requires the capability of searching through online repositories of information in an efficient manner and for libraries to be able to support the cataloguing activities about their collections to this degree of detail..

The UNIverse project [2] is developing a library system to support a virtual union catalogue. It also offers mechanisms for facilitating cataloguing activities by enabling record supply.

This activity, can be viewed in the wider context of setting up infrastructures for libraries to share information not only about their catalogues and material, in a traditional sense, but also to prepare for what can be seen as a future enhancement of their role, sharing information about digital objects. The UNIverse system is already capable of processing the whole of the retrieval process from search and locate to order and delivery of digital objects over networks.

This paper focuses on the experiences of a sub group set up within the UNIverse project to specifically test and evaluate the record sharing capabilities of the system, and collaborative cataloguing in practice. These experiences, not only as they relate to the system, but to the wider context of networked information and metadata tags for retrieval, are presented here.

The paper begins with a description of the current state of the art in regard to online learning environments, and metadata descriptions of the learning objects, which constitute the course and other relevant material, along with current practice in union catalogue assembly and maintenance. It continues with an overview of the UNIverse project and the collaborative cataloguing experiment that was conducted within it. Finally, concluding remarks about the nature of the implications upon libraries and their present and future modes

IR US 7449

BEST COPY AVAILABLE

of cataloguing activity are made.

Acknowledgements
Some of the work described here is being carried out under the EU's Telematics for Libraries Programme, project UNIverse, and EU's ACTS Programme, project GESTALT. I thank the teams of people in these projects and in particular that of the Greek Special Interest Group from UNIverse for their support.

Online learning and education materials management

The overall importance of the role of libraries in education, and moreover in distance education [3], is well-recognised [4]. Two important factors can be cited which among others contribute to their increasing participation in educational practice. On the one hand, there is the constructivist pedagogical model influencing much of present day educational thinking, and putting great emphasis on the notions of learning by discovery and exploration, and on the other the technological innovations which enable access to increasingly wider range of materials.

As has been extensively documented, [4] [5] what this means for the librarian is that the task of mediating between learner and resources becomes more imperative, and with the added pressure that they must combine elements of professional librarianship such as enquiry and research activities, with technical expertise [6]. In addition, with both remote and on campus users, they are often the primary source of instruction for students in the use of email, database querying, and other skills.

For librarians, mediating between users and resources, is but one, albeit very important, facet of their mission. They are, of course, also required to select, acquire, organise, make accessible, and preserve material. All this, while they are being subjected to enormous increases in both the numbers of users and the amount of material they can mediate access to.

One example of the increase in material, which is relevant to the education scenario, is the increasing tendency for academic institutions to consider all sorts of content production by their teaching staff as valuable commodities, and to be looking for some kind of asset management system to handle this in-house material. This content is typically primary content material, made up of lecture notes and assignments, reading lists and exam questions. But as tutors begin to explore the possibilities of new technologies for teaching, and bow to the pressure to provide content which can be transmitted to remote students, the material becomes increasingly multimedia.

Historically, either the content authors kept control of such material, or in some other cases, the computing services department, as technical experts, were given custody. However, as the volume of such material increases, and with the realisation by education service providers of the potential for exploitation of this material, the need for adequate management becomes more and more pressing. Furthermore, the philosophy of treating this material as reusable modules is increasingly prevalent. For both the educationalist and the information scientist professional, this calls into play questions of granularity. What is the smallest unit of knowledge, and what should be visible from the catalogue for that material? There is also the question of what other information about the resource should be recorded. Sufficient descriptions of the modules are required, so that they can be searched and located, and in addition displayed and manipulated. Digital resources have other descriptive needs, and more especially when they exist not in any tangible form, such as a CD or a video, but only as bits and bytes that can only be apprehended by the correct access platform of software and hardware. It is not surprising that the Library should be called upon to manage these assets, since it has amassed the most expertise in these areas.

At the present time, there is much research and effort going into the design of metadata for educational software, and into tying to pin down standards that will enable interoperablity of implemented metadata,

and particularly in regard to learning object metadata. In this respect, one can mention, the work of the Dublin Core [7], the IEEE LTSC [8], and the IMS [9] in the States, and the CEN/ISSS working group on Learning Technologies [10] in Europe and the ACTS funded GESTALT project [1]. The GESTALT project looks at the process of online learning from a holistic viewpoint, seeing the whole of the process from searching for a course, via an electronic broker, or Resource Discovery Service, to the student enrolling in a Learning Environment, to follow a programme of study and making use of assets (both primary educational content and supporting material) from the Asset Management System. GESTALT is in the process of defining metadata sets based upon the emerging standards for ensuring interoperability of the whole system. Again, in accordance with emerging standards, the encoding of the metadata will be done in XML [11]. This paper is not the place for discussion of these very interesting developments, instead, it wishes to point out the very real burden that will be placed afresh on librarians who will be asked to manage educational digital content for education service providers. For whereas professional publishers of digital material may go some way to help with pre-cataloguing items, it is doubtful whether educational content providers will do so, or will be able to do so, unaided.

For the librarian to be able to cope with the new influx, some re engineering of present modus operandi may have to be undertaken. In the next section, suggestions and solutions for addressing various parts of this complex activity are presented

Co-operation and Collaboration: Linking publishers and national bibliographies; MARC and metadata; Union catalogues and virtual union catalogues

It has been recognised by the library world that bibliographic control over electronic publications (especially those published via networks) is not adequate in the face of the continuous growth in the amount of material being published chiefly or solely in electronic form. Equally disturbing is the recognition that there is no agreed standard of bibliographic description for electronic publications. These were two of the issues that the BIBlink [12] project, funded by the EU, attempted to go some way to tackling. The BIBLink project, grew out of the CoBRa project [13] which recognised that the significant growth in electronic publishing raised issues that needed to be addressed at an international level. Project BIBLINK called upon the bibliographic expertise of the national libraries of Europe, working in conjunction with partners in the book industry, to examine ways that electronic publications are described for catalogues and other listings.

Thus BIBlink spent effort mapping from various MARC formats to various metadata schema. They found that several MARC formats were going through the process of being updated to enable cataloguing of electronic publications, in particular on-line publications. MARC format has unique value for integrating metadata describing electronic resources into existing legacy systems. If libraries wish to integrate metadata into their existing systems, and use existing software (albeit with some updating to deal with new fields) then MARC offers a solution. Indeed, most work has been done on adapting the USMARC format for the cataloguing items accessible through the Internet. OCLC's Intercat [14] project has served as a test bed for the cataloguing of network resources, and as a means to introduce and verify new fields and fine tune as required. Over 200 libraries participated in this project, the majority of them academic (60%) and nearly all of them situated in the US. There are at present nearly 83,000 records in the InterCat database.

To understand why MARC formats should be extended, it is necessary to understand something of the topology of metadata. An essential aspect of the level of richness of a format is the extent of the content, both in terms of range and depth. The attempt to describe more or less aspects of an object will be reflected in the overall level of complexity, for example designation or format rules for content. In order to identify the extent of content the elements describing an object can be clustered into groups.

An example may be seen in a reference model for business-acceptable communication proposed by

Bearman [15]. This defines clusters of data elements which would be required to fulfil the range of functions of a record. The functions of records are identified as the provision access and use rights management, networked information discovery and retrieval, registration of intellectual property, and authenticity. The clusters of data elements are defined in six layers:

1. Handle Layer
    - registration metadata or properties
    - record identifier
    - information discovery and retrieval
2. Terms and Conditions Layer
    - rights status metadata
    - access metadata
    - use metadata
    - retention metadata
3. Structural Layer
    - file identification
    - file encoding metadata
    - file rendering metadata
    - record rendering metadata
    - content structure metadata
    - source metadata
4. Contextual Layer
    - transaction content
    - responsibility
    - business function
5. Content Layer
    - content description
6. Use History Layer

From the above, it is clear that Bearman's model looks at the record in a wider context than the bibliographic context alone, and it is particularly relevant to this paper as it takes account of the business context in which metadata is used. Bearman includes metadata elements that are appropriate for metadata in the context of publishing and supply. In the new model of educational content suppliers, some of these business related metadata will be needed, if education service providers are to market their courses in a global competitive market, and if they are to deliver globally, then it is essential that the metadata take account of delivery mechanisms.

Taking the issue of cataloguing electronic resources from another angle, there have been several attempts to catalogue resources on the Internet in both automated and collaborative fashions. Take for instance, the amount of work on subject gateways [16]. Subject gateways are labour intensive to develop and maintain. They require the constant input of staff who hand pick, classify and catalogue each Internet resource. This is both the strength and the weakness of gateways. The human input allows for semantic judgements and decisions that are the key ingredient for creating a quality controlled gateway. This ingredient is lacking in automated indexes or search engines which can not filter information in such a meaningful way. However, considerable time and effort is needed to make these judgements and decisions and this means that the collection of resources is often small and slow to grow. As the number of resources available over the Internet increases, gateways need to develop ways of increasing the number of resources they can catalogue. The DESIRE project [17]. has identified two ways in which this might be done: firstly by distributed cataloguing, which increases the number of people adding resources, and secondly by automatic metadata entry: improving the efficiency of the cataloguing process.

In order to perform automatic metadata entry, subject gateways would harvest the metadata produced by subject communities into templates. One of the main issues of automatically generating templates is ensuring that the high standards (that set apart gateways from automatic search engines) are maintained. This means that both the resources included in the database should be of high quality as well as the catalogue records themselves.

The DESIRE researchers suggest that ensuring the integrity of resources could be achieved by only harvesting automatically metadata from 'trusted' information providers. A trusted provider would be a site or organisation that had been previously evaluated by the information gateway as a high quality resource. To ensure that the catalogue records remained of a high and consistent quality the gateways would need to promote 'good use guidelines' (including the use of controlled vocabularies) for the production of metadata within their subject community

Along the same lines, in September of 1998, the OCLC launched a worldwide call for participants to their Co-operative Online Resource Catalog (CORC) [18] project seeking to automate cataloguing of Internet resources. The aim of the project would be to explore the co-operative creation and sharing by libraries of metadata. Besides libraries, museums, archives, publishers and other institutions that face similar problems with the proliferation of resources on the Web are invited to participate. The project will build upon OCLC's prior activities in creating Internet resource databases through such projects as the OCLC NetFirst [19] and InterCat [20] databases, but the CORC project will rely more heavily on automated means to build its database. Both NetFirst and InterCat records will be used initially to seed the CORC database. Both full USMARC cataloguing and an enhanced Dublin Core metadata mode will be used.

As can be seen from the above two projects, fundamental to these efforts is the co-operation and collaboration of library and other staff. They have been able to build on pre-existing shared cataloguing activities to create networks that enable quicker responses to the problem of the influx of the web. These shared cataloguing activities are at the heart of this paper, and so deserve further scrutiny.

The idea of collaborative cataloguing is not new, but it was enabled by technology. From the time MARC was introduced, and libraries began the tremendous job of converting from physical card catalogues to machine readable ones, the idea of commercial record supply and union catalogues began to take hold. In the late 1960s, the convergence of technology and a good idea brought the library world into a new era of shared goals and resources. According to the OCLC, the "visionary dream" of co-operative cataloguing is now deeply embedded in library economics, and the result has been the most widely used academic database on the Internet, WorldCat (the OCLC Online Union Catalog) [21].

The step from union catalogues to virtual union catalogues has had to wait until technology was mature enough to support networking, but still there are the known problems of rights of access, etc. The best-publicised example of virtual union catalogues is that of the Virtual Canadian Union Catalogue (vCUC) [22]. The concept of the vCUC involves a decentralised, electronically accessible catalogue created by linking the databases of several institutions. The full implementation of a distributed, linked union catalogue to support all aspects of resource sharing is a complex process involving the resolution of technical, policy and service issues. Obviously, these issues cannot be tackled all at once, therefore the initiative is limited to five interlinked issues. These are: the primary use of union catalogues in support of interlibrary loan, and to identifying and resolving issues related to the record syntax to be used (USMARC and/or CAN/MARC); the provision of holdings information (accessibility and coding); the roles and responsibilities of the union catalogue participant; the standardisation of the use of library symbols; and finally, the format and degree of detail for holdings data.

For some, virtual union catalogues are still too fraught with insoluble issues to be viable. For instance, in a

nationally funded project to produce specifications for a union catalogue of university libraries in Greece [23], the decision was made to design a union catalogue with a centralised database, rather than the virtual model with distributed databases. Although this decision was not considered by all those involved to be the most forward thinking, it was seen as the most pragmatic in a region very behind in terms organised library co-operation. As their report explains, many libraries have automated systems and have processed part of their collections, but there is no shared cataloguing activity, every library does its own cataloguing independently. The only co-operation patterns to have evolved are among academic and research libraries that subscribe to a serials co-operative catalogue, operated by the National Documentation Centre. As with most countries in this situation a certain amount of leapfrogging will take place and the design the centralised catalogue of 32 higher education institutions can be seen as a first step in bringing collaborative and networking to the Greek Academic Librarian, and breaking the mould of isolation.

The case of the Greek academic and research libraries has been picked out as it provides the background for the collaborative cataloguing experiment taking place within the Greek group of libraries that is testing the UNIverse system.

UNIverse and collaborative cataloguing

Within the European funded project UNIverse a large-scale project based on the concept of a virtual union catalogue, a series of advanced library services to both end-users and librarians are offered, namely:

- Search and Retrieve - very large scale, transparent multi-database searching
- Mixed-media document delivery - integrated to the search and retrieve process
- Inter-Library Loans - integrated to the search and retrieve process
- Collaborative cataloguing/ record supply - an efficiency gain for the librarian.

The virtual union catalogue forms the core of the UNIverse system around which a number of key features have been built. Firstly, the ability to perform parallel searches upon multiple physical databases which have a variety of access methods, record syntax, character sets and languages, and see the results as if a single logical database were being searched. Secondly, the multiplicity of data sources is hidden from the user and a high quality of service is achieved both in terms of performance and data quality through record de-duplication and merging. Thirdly, through the use of Open Distributed Processing techniques the architecture has potentially unlimited scalability whilst maintaining high performance.

The libraries that are testing and validating the collaborative cataloguing aspects of the system are those in the Greek group headed by the National Library of Greece. This group comprises universities, a professional society library, and the library of an internationally renowned college. While there are some overlaps in their collections, the group's main cohesion derives from the willingness of its librarians to enter into such experiments, and their hopes that this will lead to greater collaboration between their institutions.

In the wider context, some of the aims and benefits of a Collaborative cataloguing service are a better use of staff resources; enhanced records; mutual benefit to specialist libraries; contribution to virtual union catalogue; potential source of revenue for supplying libraries. However, in the context of the Greek group of libraries, whose history of collaborative cataloguing is non existent, their hopes are more specific. Universe offers the attraction of a *virtual* Union Catalogue, with all the advantages of immediacy, flexibility, and scalability. Each institution involved employs substantial number of cataloguers as a proportion of its total staff, they hope UNIverse will offer a better use of these staff resources in terms of quicker throughput of material; substantial lessening of cataloguing backlog; better quality records. They understand the virtues of collaborative cataloguing as opposed to simple record supply, which will also enable them to share specialist subject knowledge

The libraries are at present engaged in evaluating the system. The plan is to test the use of the collaborative cataloguing scenario over five features of the record supply service. These features are: search and retrieve records for download using a variety of fields; merge records/multiple records; to create records; to enhance records; to test the use of the audit trail where libraries use the Universe Client, and the server is able to record data for the supplying library. Wherever possible each library will play both supplier and recipient roles.

Technically, the system is simple to understand. Initially, the user will search a number of targets using the UNIverse client. This search process will generate a result list that the user can select records from. When the user selects the option to download (or export) the record, a dialogue is presented to allow the file name to be specified and the required record format/syntax. Typically the local catalogue system will have a daemon process running that looks for files appearing in a pre-determined director. When new files appear the process will import the records into the local database. The record download system will then be used to place records into this directory causing them to be automatically uploaded into the local catalogue. (This daemon process is not part of the Universe system).

Some predictions for the future

MARC has been with us for nearly 30 years and has been very useful, but the new Internet and web enabled communications require new indexing paradigms, or at least extensions to existing MARC. However, the vision of embedding, or attaching, other digital information to the -- bibliographic- record is strong. The influx of digital resources is already overwhelming, the expected influx of educational material promises to place even more urgent demands upon education services providers' asset management staff. The problems are still looking for the best way to apply solutions. The technological change affects the objects to be described and the systems used to manage bibliographic data. The issue was laid out succinctly by Hickey: "Now, libraries need a system to create and share metadata for online resources to help automate resource selection, creation of the metadata itself and maintenance of links." [24] Fundamental to the technical system of creating and sharing metadata, will be the same types of human centred networks already existing for collaborative cataloguing activities. The metadata will probably exceed by far the level of detail found in the average bibliographic record. As we have tried to show in this paper, and is the experience of the UNIverse Greek SIG, collaborative cataloguing and eventually, sharing metadata, will in the end depend as much on the technology as on the co-operative networks of participants involved.

References

1. GESTALT (Getting education systems talking across Leading edge Technologies)

2. UNIverse: http://www.fdgroup.co.uk/research/universe/

3. Stork, H-G. Digital Libraries and their impact on Distance Learning: A European Perspective, IATUL News, Vol. 6. 1997, no.4 http://educate.lib.chalmers.se/IATUL/4-97.html

4. Holowachuk, D. The Role of Librarians in Distance Education, 1997 http://hollyhock.slis.ualberta.ca/598/darlene/distance.htm

5. Prestamo, A.M. Development of Web-Based Tutorials for Online Databases, 1998 http://www.library.ucsb.edu/istl/98_winter/article3.html

6. Hastings, K. Tennant, R. How to build a Digital Librarian, D-Lib Magazine, November 1996 http://www.dlib.org/dlib/november96/ucb/11hastings.html

7. http://purl.org/metadata/dublin_core/

8. http://grouper.ieee.org/groups/ltsc/index.html

9. http://www.imsproject.org/index.html

10. http://cenorm.be/isss/news/default.html

11. There are a wide variety of resources on XML, for an introduction see http://www.w3.org/XML/, and also the XML zone at http://xml-zone.com/

12. http://hosted.ukoln.ac.uk/biblink/

13. http://www.bl.uk/information/cobra.html

14. http://orc.rsch.oclc.org:6990/

15. Bearman, David and Sochats, Ken. Metadata Requirements for evidence. Available at URL: http://www.lis.pitt.edu/~nhprc/model.htm.

16. Subject gateways to be mentioned here:
DutchESS (The Dutch Electronic Subject Service, Netherlands
http://www.konbib.nl/dutchess/docs/info.html#8
EEVL (The Edinburgh Engineering Virtual Library, UK) http://www.eevl.ac.uk/volunt.html
ADAM (Arts, Design, Architecture and Media Information Gateway, UK http://adam.ac.uk/friends/
Biz/ed (Economics and Business Education Gateway, UK)
http://www.bized.ac.uk/inform/infohome.htm
EELS (Engineering Electronic Library, Sweden) http://www.ub2.lu.se/eel/about.html

17. Worsfold, E Distributed and Part-Automated Cataloguing (A DESIRE Issues Paper) March 1998 http://www.sosig.ac.uk/desire/cat/cataloguing.html and http://www.desire.org/, there is also a Desire 2 project, see

18. ftp://ftp.rsch.oclc.org/pub/Internet_cataloguing_project/Manual.txt

19. see http://www.oclc.org/oclc/new/n234/prod_netfirst_continues_growth.htm, for most recent news

20. see [14]

21. http://www.oclc.org/oclc/promo/7775os/worldcat.htm

22. http://nlc-bnc.ca/resource/vcuc/index.htm

23. This report, in Greek, but with an English language summary can be found at http://www.ntua.gr/library/deliv01.htm

24. ftp://ftp.rsch.oclc.org/pub/Internet_cataloguing_project/Manual.txt

---

Back to contents

Last edited by J.F, 31st May, 1999.